

# Dynamic Spectral Bisection

*Horst D. Simon,  
Andrew Sohn, and John Wu*

National Energy Research Scientific Computing Center  
(NERSC), Berkeley, CA 94720, [hdsimon@lbl.gov](mailto:hdsimon@lbl.gov)

October 14, 1999

# Closing the Book on Spectral Partitioning

*Horst D. Simon,  
Andrew Sohn, and John Wu*

National Energy Research Scientific Computing Center  
(NERSC), Berkeley, CA 94720, [hdsimon@lbl.gov](mailto:hdsimon@lbl.gov)

October 14, 1999

# Outline

- Overview
- Ten years of spectral mesh partitioning
- Recent results with S-HARP
- Other related work in LSI and MATLAB\*P
- Conclusions and problems for the future

# Over 10 years of mesh partitioning

- Results in the FEM community by Farhat, and by Raefsky, Nour-Omid, and Lyzenga in 1987 and 1988
- First spectral partitioning presented by Pothen and HDS at SIAM Sparse Matrix Meeting in Oregon, May 1989
- Multilevel approach presented by Barnard et al. At SIAM PP conference in Norfolk in 1992
- TMC, Centric adapt spectral in 1993 for commercial software

# Spectral Partitioning in 1994

## DESIGNING and BUILDING PARALLEL PROGRAMS

Concepts and Tools for  
Parallel Software Engineering



Ian Foster

# Developments since 1994

- Further improvements of mutlilevel approach
- Kernighan-Lin type methods improved
- excellent software packages
- we tried hard to prove that spectral approach works for dynamic load balancing (HARP, S-HARP); collaborations with Pramono, Biswas, Sohn, Olikier ...

# Summary

- Demonstrated that fully scalable implementation of spectral algorithm is possible
- Performance in the competitive range of other algorithms
- software freely available  
([www.cs.njit.edu/sohn/sharp](http://www.cs.njit.edu/sohn/sharp))
- 10 years is enough and it's time to close this chapter
- Question for the discussion: we have a great toolbox today in 1999; what else do we need to do in partitioning algorithms?

# Open Problems (Part 1) - Work to be Done

## PRACTICAL

- 1) Apply our algorithmic knowledge to real applications

## THEORETICAL

- 2) Understand the role and use of higher eigenvectors in spectral partitioning
- 3) Understand the relationship of the spectrum of the Laplacian matrix of a graph to the spectrum of the distance matrix of the graph



- Joint work with Alan Edelman (MIT) and Parry Husbands
- client server model running on the T3E and IBM SP, O2000 under development
- see me if you want to use it
- main motivation is to develop a toolbox for information retrieval AND to work with really large data sets
- SEE <http://www.nersc.gov/~parry/text/ppserver/index.html>

# MATLAB\*P in action

Parallelism through Polymorphism!!!

```
>> a=randn(512,512*p); a2=ones(512*p,512);  
m=sprand(10000,1000*p,0.01);
```

```
>> whose
```

Your variables are:

Name	Size	Bytes	Class
a	512 x 512p	1048576	ddense array
a2	512p x 512	1048576	ddense array
m	10000 x 1000p	810176	dsparse array

Grand total is 624560 elements using 2907328 bytes

```
>> b=inv(a); c=a*b; c(1:3,1:3)
```

ans =

```
1.0000 0.0000 -0.0000  
-0.0000 1.0000 0.0000  
0.0000 -0.0000 1.0000
```

```
>> e=eig(a);plot(e,'*');axis([-30 30 -30 30]);axis('square')
```

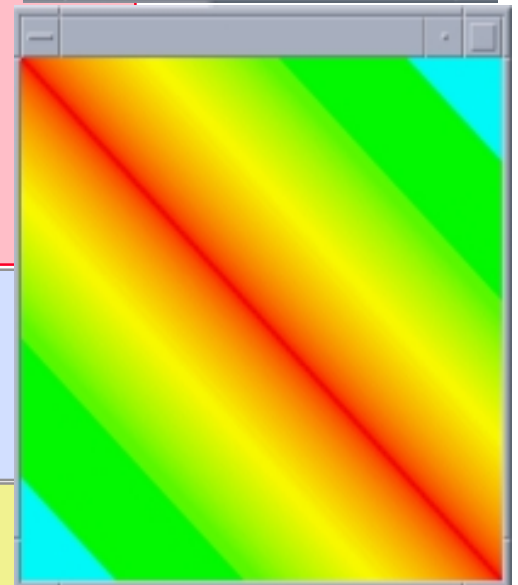
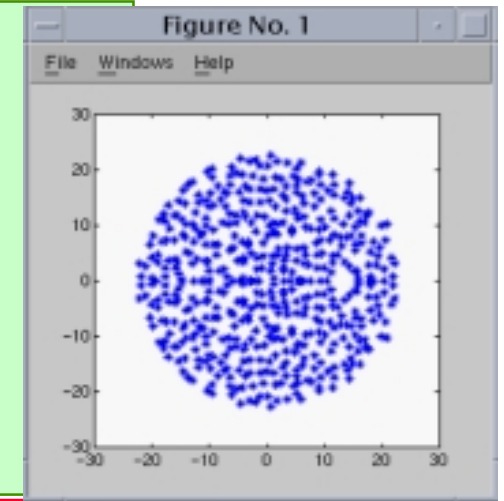
```
>> [u,s,v]=svds(m,5);s'
```

ans =

```
7.7153 7.7342 7.7447 7.7831 16.9842
```

```
>> id=eye(1000*p);x=cumsum(id,1);y=cumsum(x,1);
```

```
>> imagesc(y+y')
```



# Latent Semantic Indexing (LSI) and Information Retrieval

- LSI is equivalent to spectral partitioning
- LSI computes SVD of term-document matrix
- consider the term-term matrix (edges when two terms appear in the same document)
- LSI term vectors (singular vectors) are the same vectors as eigenvectors derived from spectral partitioning
- Current work done in Berkeley
  - understand limitations of LSI
  - sparse representation of singular vectors (Zha and Zhang)
  - see <http://www.nersc.gov/~parry/text/index.html>

# Open Problems (part 2) - new directions

## INFORMATION RETRIEVAL

- 4) Apply all the wonderful graph partitioning algorithms to information retrieval; develop a “better” LSI
  - 5) Solve the “Yahoo problem”
    - our knowledge is hierarchically organized
    - the really intriguing graph problem is how to find trees for information retrieval
- see me if you have ideas!